

FABRICATION OF GATE DIELECTRIC IN NONVOLATILE MEMORIES HAVING
SELECT, FLOATING AND CONTROL GATES

Yi Ding

5 BACKGROUND OF THE INVENTION

[0001] The present invention relates to nonvolatile memories.

[0002] Fig. 1 illustrates a flash memory cell 110 described in U.S. patent no. 6,057,575 issued May 2, 2000 to Jenq. The cell is formed in and over a semiconductor substrate 120. Silicon dioxide 130 is thermally grown on substrate 120 to a thickness of 180 Å. Select gate 140 is formed on oxide 130. Silicon dioxide 150 (100 Å) is thermally grown on a region of substrate 120 not covered by the select gate. ONO 154 (a sandwich of a layer of silicon dioxide, a layer of silicon nitride, and another layer of silicon dioxide) is formed on select gate 140. Floating gate 160 is formed on dielectric layers 150, 154. A portion of floating gate 160 overlies the select gate 140.

15 [0003] ONO layer 164 is formed on the floating and select gates. Control gate 170 is formed on ONO 164. The control gate overlies floating gate 160 and select gate 140.

[0004] N⁺ source and drain regions 174, 178 are formed in substrate 120.

[0005] Fig. 2 shows a circuit diagram of a memory array of cells 110. This is a NOR array. Each cell is shown schematically as a floating gate transistor and a select transistor connected in series. Select gate lines 140, control gate lines 170, and source lines 178 extend in the row direction (Y direction) throughout the array. Each select gate line 140 provides the select gates for one row of the array. Each control gate line 170 provides the control gates for one row. Each source line 178 is connected to source/drain regions 178 of two adjacent rows (here the same numeral 178 is used for the source lines and the source/drain regions). Bitlines 180 extend in the column direction (X direction). Each bitline 180 is connected to the regions 174 of two adjacent columns.

25 [0006] A cell 110 is programmed by hot electron injection from the cell's channel region (the P type region in substrate 120 below the cell's floating and select gates) to floating gate 160. The cell is erased by Fowler-Nordheim tunneling of electrons from

floating gate 160 to source line region 178. The cell is read by sensing a current on the corresponding bitline region 174.

[0007] The memory includes peripheral circuitry (not shown) including decoders, sense amplifiers, high voltage generators, and other circuits used for the memory access.

5 It is desirable to provide improved manufacturing techniques for fabricating the memory and the peripheral circuits.

SUMMARY

[0008] This section summarizes some features of the invention. Other features are described in the subsequent sections. The invention is defined by the appended claims
10 which are incorporated into this section by reference.

[0009] In some embodiments of the present invention, one or more peripheral transistor gates are formed from the same layer as the select gate. The gate dielectric for these peripheral transistors and the gate dielectric for the select gates (such as dielectric 130 in Fig. 1) are formed simultaneously. Elimination of separate processing steps to
15 form the gate dielectric for these peripheral transistors advantageously reduces the overall thermal processing involved in the memory fabrication.

[0010] In some embodiments, these peripheral transistors are high voltage transistors. High voltage transistors are transistors exposed to super high voltages. "Super high voltages" are voltages of a higher magnitude than the voltages provided to a memory cell
20 in a read operation. For example, suppose the read voltages are all in the range from V_{ss} to V_{cc} , where V_{ss} is ground, and V_{cc} is a power supply voltage (received from an external power supply or generated internally) of 3.0V. Then a super high voltage is a voltage of a magnitude above 3.0V.

[0011] High voltage transistors may need a fairly thick gate dielectric. Select gates
25 may also need a thick dielectric. In some embodiments, the memory is erased through the channel region. In the erase operation, the voltage between the substrate and the select gate is super high, so the select gate dielectric is made thick to sustain this voltage. The select gate dielectric matches the dielectric thickness of the high voltage transistors.

[0012] In some embodiments, the select gate dielectric is thicker than the dielectric
30 underlying the floating gates (such as dielectric 150 in Fig. 1). In some embodiments, the

memory includes peripheral transistors with different gate dielectric thicknesses, and the select gate dielectric is at least as thick as any peripheral transistor gate dielectric.

[0013] Other features and embodiments of the invention are described below. The invention is defined by the appended claims.

5 BRIEF DESCRIPTION OF THE DRAWINGS

[0014] Fig. 1 shows a vertical cross section of a prior art flash memory cell.

[0015] Fig. 2 is a circuit diagram of a prior art memory array.

[0016] Fig. 3 is a top view of a memory array according to an embodiment of the present invention.

10 [0017] Figs. 4, 5A, 5B, 6A, 6B, 7A, 7B, 8, 9, 10, 11A, 11B, 12, 13A, 13B, 14, 15A, 15B, 16, 17A, 17B 18, 19A, 19B, 20A, 20B, 21A, 21B, 22, 23A, 23B, 24A, 24B, 25A-25C show vertical cross sections of integrated circuit structures according to embodiments of the present invention.

15 [0018] Fig. 25D is a top view of an integrated circuit structure according to an embodiment of the present invention.

[0019] Fig. 25E shows a vertical cross section of an integrated circuit structure according to an embodiment of the present invention.

[0020] Fig. 25F is a top view of an integrated circuit structure according to an embodiment of the present invention.

20 DESCRIPTION OF SOME EMBODIMENTS

[0021] The embodiments described in this section illustrate but do not limit the invention. The invention is not limited to particular materials, processing steps, or dimensions. The invention is defined by the appended claims.

25 [0022] One memory embodiment of the present invention is shown in top view in Fig. 3. The memory circuit diagram is identical to that of Fig. 2. Select gate lines 140, control gate lines 170, and source lines 178 run through the memory array in the Y direction (row direction). The bitlines (not shown) run in the X direction (column

direction). The bitlines contact the corresponding source/drain regions 174 ("bitline regions") in areas 174C marked with a cross. Floating gates 160 are marked with dashed crosses. In this embodiment, the floating gates do not overlie the select gates. Control gate lines 170 overlap the select gates. Each dashed line 140E marks an edge of a select gate line 140 under a control gate line 170. Each control gate line 170 has an edge 170E1 overlying a select gate line 140, and another edge 170E2 which does not overlie the select gate line but runs at some distance D from the select gate line. The edges 170E2 and the distance D can be defined in a self-aligned manner as explained below. The edges 170E2 also define the edges of the floating gates 160 on the side of bitline regions 174. The floating gates can be completely self-aligned (i.e. defined independently of photolithographic alignment), as described below.

[0023] In Fig. 3, floating gates 160 are adjacent to bitline regions 174, not to source line regions 178 as in Fig. 1. The increased distance between the floating gates and the source lines makes it possible to increase the source line doping concentration, and thus reduce the source line resistance, because the electrons are less likely to leak from the source lines to the floating gates. Further, in some embodiments, the memory cells are erased through the channel region. The exemplary voltages are given in Table 1 below. The voltage difference between the source line region 178 and select gate 140 is fairly low (at most about 6V in Table 1, for the erase operation). Therefore, the current leakage between source line 178 and select gate 140 is low. Further, the voltage difference between the source line 178 and the substrate 120 (P well 120W) is at most a diode drop (during the erase operation), so the source line junction breakdown is unlikely. Consequently, the source line doping can be increased to reduce the sheet resistance. The invention is not limited to such embodiments however. For example, the floating gates can be adjacent to the source lines. The memory can be erased through source lines 178 or bitline regions 174.

[0024] Substrate isolation trenches 220T run through the array in the X direction. Trenches 220T are filled with dielectric 220, but dielectric 220 is etched out of the trenches at the location of source lines 178. Active areas 222 run through the array between the trenches 220T. Each active area 222 includes active areas of individual cells in one memory column. The active area of each cell consists of the cell's source/drain regions 174 and 178 and the P type channel region extending between the regions 174, 178. Numeral 178 denotes both a source line and a source/drain region ("source line

region”) of one memory cell.

[0025] Some of the figures below illustrate vertical cross sections of intermediate structures obtained during the memory fabrication. The sectional planes are indicated in Fig. 3 by lines X1-X1', X2-X2', Y1-Y1', and Y2-Y2'. The line X1-X1' runs in the X direction through an active area 222. The line X2-X2' runs in the X direction through a trench 220T. The line Y1-Y1' runs in the Y direction through a select gate line 140. The line Y2-Y2' runs in the Y direction through a control gate line 170 and floating gates 160.

[0026] In one embodiment, the memory is fabricated as follows. Substrate isolation regions 220 are formed in P doped substrate 120 by shallow trench isolation technology (“STI”). See Fig. 4 (cross section Y1-Y1'). Each region 220 is a dielectric region formed in a trench 220T. Suitable STI processes are described in U.S. patent no. 6,355,524 issued March 12, 2002 to Tuan et al.; U.S. patent application no. 10/262,785 filed October 1, 2002 by Yi Ding; and U.S. patent application no. 10/266,378 filed October 7, 2002 by C. Hsiao, all incorporated herein by reference. Other STI and non-STI processes are also possible. We will sometime refer to dielectric 220 as “STI oxide” because it is silicon dioxide in some embodiments. The invention is not limited to such embodiments or to silicon integrated circuits.

[0027] Substrate isolation regions are also formed in the memory peripheral area (not shown in Fig. 4). The peripheral area contains circuitry needed to access the memory, and may also contain unrelated circuitry (the memory may be embedded into a larger system).

[0028] As shown in Fig. 4, oxide 220 protrudes above the substrate 120. The protruding portions are shown at 220P. An exemplary thickness of portions 220P is 0.12 μm for a 0.18 μm fabrication process (a process with a 0.18 μm minimum line width). The exemplary dimensions given in this section assume a 0.18 μm fabrication process unless mentioned otherwise.

[0029] Dopant is implanted into substrate 120 to form an N type region 604 underlying the memory array. Dopant is also implanted into the substrate around the array to form a surrounding N type region (not shown) extending from the top surface of substrate 120 down to region 604. These implants create a fully isolated P well 120W for the memory array. Region 604 is not shown in the subsequent drawings, and the P well 120W is shown simply as substrate 120.

[0030] Silicon dioxide 130 (Fig. 5A, cross section Y1-Y1', and Fig. 5B, periphery) is thermally grown on the exposed areas of substrate 120 to provide gate dielectric for the select gates of the memory array and for the peripheral transistors. An exemplary thickness of oxide 130 in the array area is 120 Å. Generally, the oxide thickness depends on the maximum voltage that the oxide 130 is designed to sustain during the memory operation.

[0031] In the example shown in Fig. 5B, the peripheral area includes a high voltage transistor area 512H and a low voltage transistor area 512L. Oxide 130 is grown thermally to a thickness of 60 Å over the entire wafer. This oxide is removed from the low voltage area 512L by a masked etch. The wafer is re-oxidized to re-grow silicon dioxide in area 512L to a thickness of 60 Å. The oxide thickness in the memory array area and in high voltage area 512H increases from 60 Å to 120 Å during this step.

[0032] Thus, oxide 130 in the array area and oxide 130 in the high voltage peripheral area 512H is formed simultaneously in these two oxidation steps. The oxide 130 in area 512L and the oxide 130 in the array area and area 512H are not formed simultaneously because the oxide 130 in area 512L is formed in the second oxidation step.

[0033] As shown in Fig. 6A (cross section Y1-Y1') and Fig. 6B (periphery), intrinsic polysilicon layer 140 is formed over the structure by a conformal deposition process (e.g. low pressure chemical vapor deposition, "LPCVD"). Polysilicon 140 fills the spaces between the oxide protrusions 220P in the memory array area. The top polysilicon surface is planar because the polysilicon portions deposited on the sidewalls of protrusions 220P meet together.

[0034] Fig. 6B may represent either the low voltage or the high voltage transistor area. In some embodiments, there are more than two peripheral areas with different gate oxide thicknesses, and Fig. 6B may represent any of these areas.

[0035] Polysilicon 140 covers the regions 120i (Fig. 6B) at the interface between substrate 120 and field oxide 220 in the peripheral area. Polysilicon 140 will protect the oxide 220 in this area to prevent formation of grooves ("divots") during subsequent processing. Polysilicon 140 will be used to form the peripheral transistor gates. The grooving in regions 120i under the transistor gates is undesirable because it degrades the transistor characteristics.

[0036] Non-conformal deposition processes, whether known or to be invented, can also be used for layer 140. If the top surface of polysilicon 140 is not planar, it is believed that the polysilicon 140 can be planarized using known techniques (e.g. CMP, or spinning a photoresist layer over the polysilicon 140 and then simultaneously etching the resist and the polysilicon at equal etch rates until all of the photoresist is removed). The bottom surface of polysilicon 140 is non-planar as it goes up and down over the oxide protrusions 220P.

[0037] An exemplary final thickness of polysilicon 140 is 0.16 μm over the active areas.

10 [0038] Silicon dioxide layer 780 (Fig. 6B) is formed over the wafer, by TEOS CVD for example, to a thickness of 400-500Å. This layer will serve as an etch stop in a silicon nitride etch. Optionally, oxide 780 is removed from the array area by a masked etch.

[0039] The peripheral area is masked, and polysilicon 140 is doped N+ in the array area. Polysilicon 140 remains undoped ("INTR", i.e. intrinsic) in the periphery. The peripheral transistor gates will be doped later, with the NMOS gates doped N+ and the PMOS gates P+, to fabricate surface channel transistors in the periphery with appropriate threshold voltages. The invention is not limited to the surface channel transistors or any peripheral processing. In particular, entire polysilicon 140 can be doped N+ or P+ after the deposition or in situ.

20 [0040] Silicon nitride 810 is deposited on polysilicon 140, by LPCVD for example, to an exemplary thickness of 1500Å. If desired, a pad oxide layer (not shown) can be formed on polysilicon 140 before the nitride deposition. The pad oxide layer will provide an additional protection for the select gates during the patterning of control gate polysilicon 170 described below.

25 [0041] In some embodiments, the top surface of polysilicon 140 and/or nitride 810 is not planar.

[0042] The wafer is coated with a photoresist layer 820. See Fig. 7A, cross section X1-X1', and Fig. 7B, periphery. (Fig. 7B shows only the active area, not the field oxide 220.) Resist 820 is patterned to define the select gate lines 140. The peripheral area is covered by the resist. Edges 140E of select gate lines 140 are adjacent to the future positions of source lines 178. The memory array geometry is not sensitive to a

misalignment between mask 820 and the mask defining the isolation trenches 220T (Fig. 3) except possibly at the boundary of the memory array.

[0043] Silicon nitride 810 is etched through the resist openings. The resist is removed, and polysilicon 140 is etched away where exposed by nitride 810. Then the exposed oxide 130 is removed. The select gate lines are formed as a result. (In an alternative embodiment, the resist defining the nitride 810 is removed after the etch of polysilicon 140 and/or oxide 130.)

[0044] As shown in Fig. 8 (cross section X1-X1'), the structure is oxidized to grow silicon dioxide 150 on substrate 120 and the sidewalls of polysilicon gates 140 in the array area. Oxide 150 will serve as tunnel oxide on substrate 120, and will provide sidewall insulation for the select gates. The oxide thickness depends on the dopants and dopant concentrations. In one embodiment, oxide 150 is 90 Å thick on substrate 120, and is 300 Å thick on the select gate sidewalls. The peripheral area is covered by nitride 810 (Fig. 6B), and remains substantially unchanged during this step.

[0045] Floating gate polysilicon 160 (Fig. 9, cross section X1-X1') is deposited over the structure, by LPCVD for example, and is doped during or after the deposition. Polysilicon 160 is sufficiently thick to ensure that its top surface is at least as high throughout the wafer as the top surface of nitride 810. In the embodiment of Fig. 9, the top surface of layer 160 is planar due to a conformal deposition to a thickness larger than half the distance between the adjacent select gate lines 140. In one embodiment, the distance between select gate lines 140 over the future positions of bitline regions 174 is 0.8 μm, and the polysilicon 160 is more than 0.4 μm thick.

[0046] If the top surface of polysilicon 160 is not planar, it is planarized by CMP or a suitable etch.

[0047] After planarization (if needed), layer 160 is etched down without a mask. The etch end point is when STI oxide 220 becomes exposed. Fig. 10 (cross section X1-X1') shows an intermediate stage in this etch, when nitride 810 becomes exposed. At this stage, layer 160 has been removed from the periphery, so the periphery becomes as in Fig. 6B.

[0048] Figs. 11A (cross section X1-X1') and 11B (cross section Y2-Y2') show the array area at the end of the polysilicon etch. The polysilicon has been removed from the

top surface of oxide 220. In some embodiments, the final thickness of layer 160 is 1200Å. The etch is selective to nitride 810.

[0049] Optionally, a timed etch of oxide 220 is performed to recess the top surface of oxide 220 below the surface of polysilicon 160. See Fig. 12 (cross section Y2-Y2'). This etch will improve the capacitive coupling between the floating and control gates. See the
5 aforementioned U.S. patent no. 6,355,524. In the embodiment of Fig. 12, the oxide 220 continues to protrude above the top surface of substrate 120 by at least 0.10 µm. In other embodiments, the oxide 220 does not protrude above the substrate after the etch (the top surface of layer 220 is level with the top surface of the substrate after the oxide etch).

10 [0050] ONO layer 164 (Fig. 13A, cross section X1-X1', and Fig. 13B, periphery) is formed over the structure. Control gate polysilicon layer 170 is deposited on ONO 164 and is doped during or after the deposition.

[0051] The top surface of polysilicon 170 is not planar in the array area. Layer 170 has protrusions 170.1 over the select gate lines 140. Cavities 170C form in layer 170
15 between protrusions 170.1 over the future positions of bitline regions 174. The protrusions 170.1 will be used to define the overlap between the floating and control gates without additional dependence on photolithographic alignment.

[0052] In Fig. 13A, polysilicon 170 is substantially planar over the future positions of source lines 178 because the source lines 178 are fairly narrow (0.22µm width in some
20 embodiments) and layer 170 is relatively thick (e.g. 0.18µm). In other embodiments, the layer 170 is not planar over the source lines 178, and a cavity 170C forms over each source line. The topography of layer 170 depends on the underlying topography, the thickness of polysilicon 170, and the polysilicon deposition process.

[0053] As shown in Figs. 14 (cross section X1-X1'), a layer 1710 is deposited over
25 the structure and etched without a mask to expose the polysilicon 170. Layer 1710 fills the cavities 170C. When layer 1710 is etched in the array area, layer 1710 is removed in the periphery, so the periphery becomes as in Fig. 13B. In one embodiment, layer 1710 is silicon nitride deposited to have a planar top surface or planarized during the etch.

[0054] In some embodiments, the etch of nitride 1710 continues after the exposure of
30 polysilicon 170, and the nitride etch exposes the sidewalls of polysilicon protrusions 170.1 (Fig. 13A). Whether or not the polysilicon sidewalls are exposed, the exposed

edges of polysilicon 170 define the control gate edges 170E2 (Fig. 3) as described below. Therefore, the edges 170E2 and the distance D are defined without resort to photolithography. In some embodiments, $D=0.18\mu\text{m}$. The overlap between the floating and control gates is also defined without photolithography.

5 [0055] The wafer is oxidized to grow silicon dioxide 1720 on the exposed polysilicon 170. See Fig. 15A (cross section X1-X1') and Fig. 15B (periphery). An exemplary thickness of oxide 1720 is 500\AA .

[0056] In some embodiments, layer 1720 is some other material formed selectively on polysilicon 170. For example, layer 1720 can be a conductive metal silicide formed by
10 a salicide (self-aligned silicidation) technique.

[0057] The wafer is coated with photoresist 1730 (Fig. 16, cross section X1-X1'). Openings are formed in the resist over the future positions of source lines 178. The location of the longitudinal edges of mask 1730 is the location of the future positions of control gate edges 170E1 (see also Fig. 3). These edges can be located anywhere over
15 select gate lines 140. The resist is removed from the peripheral area.

[0058] Oxide 1720 and at least a portion of polysilicon 170 are removed where exposed by resist 1730. See Fig. 17A, cross section X1-X1', and Fig. 17B, periphery. The etch of polysilicon 170 may stop when ONO 164 is exposed, or may continue after the exposure of ONO 164. In either case, polysilicon 170 is etched away in the periphery.
20 When ONO 164 is exposed, the etch may continue for a predetermined time (a timed etch), or may continue until all of the exposed polysilicon 170 is removed. In one embodiment, the polysilicon etch is a timed etch reducing the thickness of polysilicon 170 over the source lines to about $0.18\mu\text{m}$.

[0059] Resist 1730 and nitride 1710 are removed. The resulting structure is shown in
25 Figs. 18 (cross section X1-X1'). The periphery remains as in Fig. 17B.

[0060] Polysilicon 170, ONO 164, and polysilicon 160 are etched with oxide 1720 as a mask. The resulting structure is shown in Fig. 19A (cross section X1-X1') and Fig. 19B (periphery). In some embodiments, the polysilicon etch of layers 170, 160 is anisotropic, and the etch of ONO 164 is isotropic or anisotropic. The etch of ONO 164 may remove
30 portions of oxide 1720 and/or nitride 810, and may also remove some oxide 150 on the sidewalls of select gate lines 140.

[0061] The wafer is coated with photoresist 2620 (Fig. 20A, cross section X1-X1'). The resist is patterned to expose the source lines 178. Each source line 178 traverses the memory array between two adjacent control gate lines 170, and provides one source/drain region to each cell in the two rows associated with the two control gate lines. The edges of the resist openings can be positioned anywhere over select gate lines 140 or floating gates 160. The periphery is covered by the resist.

[0062] Silicon dioxide 220 is etched out of trenches 220T in the areas exposed by resist mask 2620 (Fig. 20B, cross section X2-X2'). This etch removes oxide 150 in the active areas over the source lines (Fig. 20A). This etch may also remove the exposed portions oxide 1720 if oxide 1720 is not entirely covered by the resist. Then the source line implant (N+) is performed using the same mask. In some embodiments, this is a high energy, high dose implant, possibly preceded by a lower energy, low dose, large angled implant (the angle can be 10° to 30° for example), to achieve a 0.1μm to 0.2μm source line diffusion depth.

[0063] In an alternative embodiment, when the resist mask 2620 has been formed, a high energy N+ implant is performed before the etch of oxide 220, then oxide 220 is etched out of the trenches using the same mask, and then another, lower energy N type implant is performed using the same mask. The first (high energy) implant is at least partially blocked by oxide 220 in the trenches to avoid shorting the source lines 178 to N type isolation region 604 (Fig. 4). See the aforementioned U.S. patent no. 6,355,524.

[0064] Resist 2620 is removed. Another photoresist layer (not shown) is formed over the wafer and patterned to cover the array but expose the entire periphery. Then nitride 810 (Fig. 19B) is etched away from the peripheral area. Oxide 780 serves as an etch stop during the nitride etch. Then oxide 780 is removed.

[0065] The resist covering the array is removed, and another photoresist layer (not shown) is formed to cover the array and define the peripheral transistor gates. Polysilicon 140 is etched away where exposed by this resist.

[0066] The resist is removed. The wafer is coated with a photoresist layer 2720 (Fig. 21B, periphery). The resist is patterned to expose the entire array area (Fig. 21A, cross section X1-X1') and also to expose the peripheral NMOS transistor regions. Fig. 21B shows a peripheral NMOS transistor region 512N with a P well 2724P, and a peripheral

PMOS transistor region 512P with an N well 2724N. These wells were defined before formation of oxide 130. There can be many regions 512N, 512P in the integrated circuit. Resist 2720 covers the PMOS transistor regions 512P. An N type implant (N-) is performed to form the LDD (lightly doped drain) extensions for peripheral NMOS source/drain regions 2730N (Fig. 21B). This implant also dopes the NMOS gates 140 in the periphery. In addition, the implant dopes bitline regions 174 (Fig. 21A) and increases the dopant concentration in source lines 178.

[0067] In some embodiments, the memory array is not exposed by resist 2720, and no doping is performed in the source lines and the bitline regions at this step.

10 [0068] Resist 2720 is removed, and another photoresist layer 2820 (Fig. 22, periphery) is formed to cover the NMOS peripheral transistor regions 512N and the memory array. A P type implant (P-) is performed to form the LDD extensions for PMOS source/drain regions 2730P and to dope the peripheral PMOS transistor gates.

15 [0069] Resist 2820 is removed. A thin silicon dioxide layer 2904 (see Fig. 23A, cross section X1-X1', and Fig. 23B, periphery) is grown on the exposed silicon surfaces of layers 140, 160, 170 by a rapid thermal oxidation process (RTO). Alternative techniques can also be used such as chemical vapor deposition (e.g. TEOS CVD), a high temperature oxide process (HTO), or other suitable techniques, known or to be invented. These techniques may form the oxide 2904 over the entire structure and not only on the silicon surfaces. An exemplary thickness of oxide 2904 is 100 Å.

20 [0070] A thin silicon nitride layer 2910 is deposited and etched anisotropically without a mask to form sidewall spacers over the gate structures. The etch of nitride 2910 may remove some of nitride 810 in the array area (Fig. 23A). If oxide 2904 was deposited over the entire structure (by TEOS CVD or HTO for example), oxide 2904 will help protect the substrate 120 during the nitride etch. Spacers 2910 meet over the source lines 178 and create a thick nitride layer over the source lines. In other embodiments, the spacers do not meet over the source lines.

25 [0071] Then N+ and P+ implants are performed to create source/drain structures for the peripheral transistors and the bitline regions 174. More particularly, the peripheral PMOS transistor area 512P is masked with resist (not shown), and an N+ implant is performed to create the source/drain structures for bitline regions 174 and the peripheral

NMOS transistors and increase the dopant concentration in the peripheral NMOS gates 140. The floating, control and select gates and the overlying nitride layers mask this implant so no additional masking in the array area is needed.

[0072] The resist is removed. The array and the peripheral NMOS transistor regions 512N are masked with a resist (not shown), and a P+ implant is performed to create the source/drain structures for the peripheral PMOS transistors and increases the dopant concentration in the PMOS transistor gates 140.

[0073] The resist is removed. A silicon dioxide etch is performed to remove the oxide 1720 and expose the control gate lines 170 (Fig. 24A, cross section X1-X1'). This etch also removes the exposed portions of oxide 150 over bitline regions 174 in the array area, the exposed oxide 130 over source/drain regions 2730N, 2730P in the periphery (see Fig. 24B), and the oxide 2904 over the peripheral transistor gates.

[0074] A conductive metal silicide layer 2920 is formed by a self-aligned silicidation (salicide) process on the exposed silicon surfaces of control gate lines 170, bitline regions 174, peripheral transistor gates 140 and peripheral source/drain regions 2730N, 2730P. The salicide process involves depositing a metal layer, heating the structure to react the metal with the silicon, and removing the unreacted metal. This can be followed by an anneal or any other suitable processing, known or to be invented, to improve the silicide properties (e.g. increase its conductivity). Titanium, cobalt, nickel, and other conductive materials, known or to be invented, can be used for the metal layer. Non-salicide selective deposition techniques, known or to be invented, that selectively form a conductive layer 2920 on the exposed silicon but not on a non-silicon surface, can also be used.

[0075] As noted above in connection with Fig. 15, layer 1720 can be a conductive metal silicide formed by a salicide process. In this case, layer 1720 does not have to be removed. The silicidation process of Fig. 24A will silicide the bitline regions 174, the peripheral gates 140 and the peripheral source/drain regions 2730.

[0076] As shown in Fig. 25A (cross section X1-X1') and Figs. 25B and 25C (periphery), inter-level dielectric 3204 is deposited over the wafer. Fig. 25C shows only an NMOS transistor region, but the PMOS regions are similar. See also Fig. 25E showing an array cross section X3-X3' described below in connection with Fig. 25F. Contact openings are etched in dielectric 3204 to expose the silicided surfaces of bitline regions

174 (Fig. 25A), source/drain regions 2730P and 2730N (Fig. 25B), peripheral gates 140 (Fig. 25C), and control gates 170 (Fig. 25E). The silicide 2920 protects the bitline regions 174 and the source/drain regions 2730 during this etch. A conductive layer 3210 (e.g. metal) is deposited and patterned to form the bitlines 180 and possibly other features. The figures also show an optional metal layer 3220 (e.g. tungsten) used to fill the contact openings before the deposition of layer 3210.

[0077] Fig. 25D (top view) shows an extension of a peripheral transistor gate 140 over STI oxide 220. The extension can be made to form a contact to the gate or for some other reason (e.g. to connect the gate to other features). The region 120i at the interface between the substrate 120 and field oxide 220 is protected from the divot formation because the gate is formed using the first polysilicon layer 140. See also Fig. 6B. The transistor of Fig. 25D can be a high voltage transistor (in area 512H in Fig. 5B) or a low voltage transistor (in area 512L).

[0078] Figs. 25E, 25F illustrate the boundary of the memory array. Contacts to control gate lines 170 and select gate lines 140 are formed in this area. Fig. 25F is a top view, and Fig. 25E illustrates a vertical cross section along the line X3-X3' in Fig. 25F. The line X3-X3' passes through control gate contact opening 170CT formed in dielectric 3204. Control gate contact opening 170CT and select gate contact opening 140C are formed over STI oxide 220. Control gate line 170 has a widened portion 170X to accommodate the contact opening 170CT. Select gate line 140 has a widened portion 140X1 to accommodate the select gate contact opening 140C.

[0079] Select gate line 140 has another widened portion 140X2 under the widened portion 170X of the control gate line. The portion 170X is created in a self-aligned manner by the widened portion 140X2. As shown in Figs. 3, 14, and 19A, the control gate edge 170E2 follows the select gate edge 140E at the distance D from the select gate. The distance D is defined without photolithography as explained above. The select gate edges are defined by mask 820 (Fig. 7A). The select gate edges are straight edges in this embodiment, but in the area shown in Fig. 25F the edge 140E deviates from the straight line to widen the select gate to form the region 140X2. Consequently, the control gate edge 170E2 deviates from the straight line to form the widened region 170X in a self-aligned manner.

[0080] Other details of the memory fabrication process for one embodiment are given

in U.S. patent application no. 10/393,212 "NONVOLATILE MEMORIES AND METHODS OF FABRICATION" filed March 19, 2003 by Yi Ding and incorporated herein by reference.

[0081] In one embodiment, the memory cells 110 are programmed by channel hot electron injection. The corresponding select gate 140 is held at a voltage sufficiently high to invert the underlying portion of the cell's channel region. Control gate 170 is driven high relative to substrate 120 to raise the voltage on floating gate 160 relative to the channel region and invert the channel region under the floating gate. A voltage difference is provided between the source/drain regions 174, 178 to induce a current and cause the hot electron injection from the channel region into the floating gate. The cells are erased by Fowler-Nordheim tunneling through the channel regions ("bulk erase"). The cells are read by sensing a current on bitlines 180 when the select gate 140 is at a high enough voltage to invert the underlying portion of the channel region, the control gate 170 is at an appropriate voltage to invert the underlying portion of the channel region if, and only if, the cell is erased, and a voltage difference is induced between the source/drain regions 174, 178. Exemplary voltages are shown below in Table 1. Vcc is assumed to be 2.7V to 3.6V. "Selected" means the memory cell is selected by the address signals. Of note, a select gate line, a control gate line, or other lines can be shared by both selected and unselected memory cells. In such cases, the "selected" voltages apply.

[0082] TABLE 1

	SG 140	CG 170	BL 180	SL 178	P well 120W
<u>Read</u>					
Selected:	Vcc	Vcc	1.0V	0V	0V
Not selected:	0V	0V	0V	0V	0V
<u>Program</u>					
Selected:	2.0V	10.0V	6V	0V	0V
Not selected:	0V	0V	Vcc	0V	0V
<u>Erase:</u>	2.0V	-10.0V	Float	Float	8V

[0083] The invention is not limited to any particular read, erase or programming techniques, to NOR memory arrays, LDD structures, to a particular array architecture or fabrication method, or to particular voltages. For example, the memory can be powered

by multiple power supply voltages. Floating gates 160 (Fig. 3) can be defined using a masked etch, and can extend over sidewalls of select gate lines 140. See U.S. patent application no. 10/411,813 filed by Yi Ding on April 10, 2003 and incorporated herein by reference. The source lines can be formed from a layer overlying the substrate 120 and
5 contacting the source line substrate regions 178; the source lines do not have to go up and down the isolation trenches 220T. Also, substrate isolation regions 220 do not have to traverse the entire array. The invention is applicable to non-flash memories (e.g. non-flash EEPROMs) and to multi-level memory cells (such a cell can store multiple bits of information). Other embodiments and variations are within the scope of the invention, as
10 defined by the appended claims.